

Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Harzing, Anne-Wil ORCID logo ORCID: <https://orcid.org/0000-0003-1509-3003> (2019) Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics*, 120 (1) . pp. 341-349. ISSN 0138-9130 [Article] (doi:10.1007/s11192-019-03114-y)

Final accepted version (with author's formatting)

This version is available at: <https://eprints.mdx.ac.uk/26508/>

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

Two new kids on the block:

How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science?

Anne-Wil Harzing

Version April 2019

Accepted for *Scientometrics*

**Copyright © 2019 Anne-Wil Harzing
All rights reserved.**

Prof. Anne-Wil Harzing
Middlesex University
The Burroughs, Hendon
London NW4 4BT
Email: anne@harzing.com
Web: www.harzing.com

Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science?

ANNE-WIL HARZING
Middlesex University
The Burroughs, Hendon, London NW4 4BT
Email: anne@harzing.com
Web: www.harzing.com

Abstract

In the last three years, several new (free) sources for academic publication and citation data have joined the now well-established Google Scholar, complementing the two traditional commercial data sources: Scopus and the Web of Science. The most important of these new data sources are Microsoft Academic (2016), Crossref (2017) and Dimensions (2018). Whereas Microsoft Academic has received some attention from the bibliometric community, there are as yet very few studies that have investigated the coverage of Crossref or Dimensions. To address this gap, this brief letter assesses Crossref and Dimensions coverage in comparison to Google Scholar, Microsoft Academic, Scopus and the Web of Science through a detailed investigation of the full publication and citation record of a single academic, as well as six top journals in Business & Economics.

Overall, this first small-scale study suggests that, when compared to Scopus and the Web of Science, Crossref and Dimensions have a similar or better coverage for both publications and citations, but a substantively lower coverage than Google Scholar and Microsoft Academic. If our findings can be confirmed by larger-scale studies, Crossref and Dimensions might serve as good alternatives to Scopus and the Web of Science for both literature reviews and citation analysis. However, Google Scholar and Microsoft Academic maintain their position as the most comprehensive free sources for publication and citation data.

Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science?

Introduction

There is a comprehensive literature comparing academic publication and/or citation coverage for the Web of Science, Scopus, and Google Scholar (for recent examples see e.g. Harzing & Alakangas, 2016; Martin-Martin, Orduna-Malea, Thelwall, & Delgado-López-Cózar, 2018), as well as a growing number of studies on Microsoft Academic (see e.g. Harzing & Alakangas, 2017a/b; Hug & Brandle, 2017; Hug, Ochsner & Brandle, 2017; Thelwall, 2017, 2018a). To date, however, very few studies have investigated coverage of the two latest new sources for academic publication and citation data: Crossref and Dimensions.

Crossref, a not-for-profit organization, was founded in 2000 by 12 publishers to simplify the process of linking to research on other publisher platforms; since then it has grown to over 11,000 members from 128 countries (Fairhurst, 2019). It has developed a wide range of functions over the years, but for this article our main interest is the addition of open citation data in April 2017, making it possible to use Crossref for citation analysis through an API. Since November 2017, Publish or Perish (Harzing, 2007) has provided the option of searching for authors, journals and key words in Crossref. Although several articles have been published on the Crossref initiative, to the best of our knowledge there have been no articles reviewing its publication and citation coverage.

Dimensions was launched by Digital Science in January 2018 (see Orduña-Malea & Delgado-López-Cózar, 2018 for an excellent summary of its history and functionality). This article focuses on the free version of Dimensions, which offers access to a subset of the data available in Dimensions Plus. To the best of our knowledge, there are only two published studies that have investigated Dimensions coverage. Thelwall (2018b) showed that for publications in the field of Food Science between 2008 and 2018 and a random sample of 10,000 publications for 2012 from all fields, coverage and citation counts of Dimensions were comparable to those of Scopus. Orduña-Malea & Delgado-López-Cózar (2018) showed that for most of 17 Library and Information Science journals, the h5 index in Dimensions was only slightly lower than for Scopus, but substantially lower than for Google Scholar. A detailed comparison for the *Journal of Informetrics* showed that both publication and citation counts for the years 2013, 2014 and 2015 were almost identical for Dimensions and Scopus. Finally, a comparison between Scopus, Google Scholar Citations and Dimensions for all 28 authors who won the Derek de Solla Price showed that citations were significantly lower in Dimensions than in both Scopus and Google Scholar Citations, thus signaling a disappointing coverage for author searchesⁱ.

This article focuses on a detailed comparison across six data sources for an academic's full publication and citation record as well as journal searches for six of the top journals in the field of Business & Economics. As such it provides three unique contributions. First, it presents the first study of Crossref coverage and allows us to verify whether, just over a year after its launch, Dimensions author coverage has improved. Second, it compares both Crossref and Dimensions coverage with no less than five other data sources. Third, by comparing author and journal coverage at the level of individual publications, it provides a more fine-grained comparison of coverage. Rather than investigating broad patterns, it thus demonstrate how an individual researcher can benefit from access to these two new data sources.

Data collection

To investigate Crossref and Dimensions coverage, I first conducted a detailed analysis of my own publication record. Despite this being a small-scale test, as discussed in more detail in Harzing (2016) there are four reasons why my own publication record is appropriate for this purpose. It includes a large number of publications in a wide range of journals in Management, International Business and Library & Information Systems, covers a 25-year period, and includes a significant variety of non-traditional publications. Finally, given that Google Scholar covers virtually all of my significant academic publications, it presents an excellent baseline for our comparison across the six data sources. Second, I compared coverage for six top journals in the field of Business & Economics, focusing on a single volume published 10 years ago, thus allowing sufficient time for citations to accrue. Within Business, I included the sub-disciplines of Management, International Business, Accounting, Finance, and Marketingⁱⁱ.

All data were collected in the second week of April 2019 with the aid of Publish or Perish (2007). Publish or Perish currently allows for searches in six data sources – Crossref, Google Scholar, Google Scholar Profiles, Microsoft Academic, Scopus and the Web of Science – and has recently implemented experimental in-house Dimensions support. Data for my own publication record were retrieved with an author search, paying special attention to the different search syntaxes in the various data sources. For Google Scholar, I used my manually curated Google Scholar Profile rather than the raw Google Scholar data. Data for the six journals were retrieved with a search for either the full journal title or an ISSN search. Only journal publications with substantive academic content were used for the comparison, thus excluding book reviews, calls for papers, editorial board notices, and errata. Results for all six data sources were subsequently exported to Excel, which allowed one-on-one matching of publications and a comparison of citation counts.

Results

Publication coverage for an individual academic record

My full publication record is comprised of 84 journal articles, four books, twenty-five book chapters, a software program, and an online compilation of journal rankings (see Table 1). It also includes more than 100 conference papers and well over 200 “other” publications, such as white papers, newsletter/magazine articles, and blog posts. However, the conference papers are generally not available online and the other publications would not typically be seen as academic publications. Hence substantive coverage of these two publication categories would not be expected in any of the six data sources.

Table 1: Publication coverage for academic record across six data sources

Data source/ Document type	Journal Articles	Books	Book Chapters	Conference papers	Other publications	Software /Data	Total
All publications	84	4	25	100+	200+	2	400+
Crossref	83	1	8	8	0	0	100
Dimensions	83	1	1	8	0	0	93
Google Scholar	84	4	24	15	19	2	148
Microsoft Academic	84	4	10	11	6	1	116
Scopus	79	0	2	2	0	0	83
Web of Science	61	1	8	0	0	0	70

As Table 1 shows, both Google Scholar and Microsoft Academic find *all* of my journal articles and books. Crossref and Dimensions miss one 1996 journal article in a journal that, after a change of publisher, no longer has online coverage for its early years of publication. Both cov-

er only one of the four books, a 2018 Palgrave publication. Scopus does not record any of the books, but *does* include most of the journal articles. The same 1996 article is missing, as are two articles in European journals that were not yet listed in Scopus in the year of publication [2003 and 2008 respectively]. The final missing article is one published in “online first” and not yet allocated to a journal issue. The Web of Science does record the 2018 Palgrave book, but covers only 61 out of the 84 journal articles.

With regard to book chapters, Google Scholar finds virtually all my book chapters, whereas Crossref, Microsoft Academic and the Web of Science find the seven individual chapters of the 2018 Palgrave book, as well as one (Crossref, Web of Science) to three (Microsoft Academic) other chapters. Dimensions and Scopus only find an incidental chapter or two. With regard to conference papers, CrossRef and Dimensions find all my Academy of Management (AoM) Proceedings papers since 2008, two of which were published in full and six with their abstracts only. Microsoft Academic and Google Scholar find the same AoM proceedings papers, but papers from quite a few other conferences too. Scopus only finds the two AoM proceedings papers that have been published in full, whereas the Web of Science finds none.

Only Google Scholar and Microsoft Academic find any of the “other” publications, mainly recording white papers, blogposts on the LSE Impact blog, and magazine articles. Further, even though it is missing the Journal Quality List, Microsoft Academic does record the Publish or Perish software. None of the four other data sources - Crossref, Dimensions, Scopus or the Web of Science – finds any of the non-traditional publications.

Overall, Crossref and Dimensions thus have a better coverage than Scopus and the Web of Science for journal articles, book chapters and conference papers and an equally poor coverage for books and non-traditional publications. However, coverage for the two other free data sources, Google Scholar and Microsoft Academic, is substantially better.

Citation coverage for an individual academic record

Citation coverage for my full publication record was monitored on a monthly basis between early December 2018 and early April 2019. In this period, citation levels for all six data sources increased at a rate of 0.5% to 1.5% per month. Differences between the data sources remained stable; thus we discuss only the most recent round of data collection.

For my full publication record, Crossref and Dimensions report citation levels of 34% and 37% of those of Google Scholar (see Table 2). This is quite similar to Scopus (38%), whereas Microsoft Academic displays a much higher (88%) and the Web of Science a much lower number of citations (23%). When only comparing journal articles, citation levels across the six data sources diverge less, with Microsoft Academic citation levels (98%) being virtually on par with Google’s, Crossref, Dimensions and Scopus sitting at 42-46% of Google Scholar citations and the Web of Science closing the ranks with 28%.

Table 2: Citation coverage for academic record across six data sources

	Crossref	Dimensions	Google Scholar	Microsoft Academic	Scopus	Web of Science
All publications	5419	5815	15871	13890	5952	3631
% of Google Scholar citations	34%	37%	100%	88%	38%	23%
Journal articles only	5415	5809	12767	12465	5936	3631
% of Google Scholar citations	42%	46%	100%	98%	46%	28%

As overall citation levels can hide large differences for individual publications, we also compared Crossref and Dimensions citations with the four other data sources for *each* individual publication. Not surprisingly given the difference in overall citation levels, both Crossref and Dimensions had lower citation levels than Google Scholar and Microsoft Academic for each of the publications.

For Scopus, the results are mixed. Although well over half my publications in Crossref and two thirds of my publications in Dimensions are within a range of -5/+5 citations when compared to Scopus, between 20% (Dimensions) and 30% (Crossref) of my publications have substantively fewer citations in these sources than in Scopus. This might well be caused by the fact that Elsevier, the provider of Scopus, doesn't support the Open Citations Initiative. At the other end of the spectrum, around 15% of my publications have substantively more citations in Crossref and Dimensions than in Scopus. However, the two most important of these are journal articles that are not included in Scopus.

Less than 20% of the publications had a lower level of citations in Crossref and Dimensions than in the Web of Science; in most cases this was a difference of only a few citations. In contrast, well over half of my publications had *more* citations in Crossref than in the Web of Science, whereas this was the case for nearly two thirds of my publications in Dimensions. There were only two articles, one in JASIST and one in Scientometrics that had a substantially higher number of citations in the Web of Science than in Crossref/Dimensions: 169 vs. 155/144 and 40 vs. 34/31.

Overall, Crossref and Dimensions thus have citation levels that are fairly similar to Scopus for most publications, substantially higher than the Web of Science for most publications, but significantly lower than Google Scholar and Microsoft Academic for all publications.

Analysis of publication and citation coverage for six Business & Economics journals

Table 3 reports publication coverage for six top journals in the field of Business & Economics. Four of the six data sources – Dimensions, Google Scholar, Scopus and the Web of Science – show identical coverage for each of the six journals. Crossref reports a higher number of publications for two of the journals, published respectively by Springer and Oxford University Press. This is caused by the fact that, for these two publishers, Crossref records the year in which the article appeared in online first as the publication year. Microsoft Academic suffers from the same problem for the OUP journal, but not for the Springer journal. However, Microsoft Academic reports far fewer articles for the Journal of Finance than the other data sources. This was caused by Microsoft Academic listing the year that the missing articles were published in the NBER working paper series as the year of publication. Oftentimes this occurred many years before the official journal publication and thus these articles were missing when the search was confined to the year 2009. These two issues were reported to Crossref and Microsoft Academic and should be relatively easy to resolve.

Table 3: Publication coverage for 2009 volume of six top journals across six data sources

Journal	Crossref	Dimensions	Google Scholar	Microsoft Academic	Scopus	Web of Science
Academy of Management Journal	63	63	63	63	63	63
Journal of Accounting Research	40	40	40	40	40	40
Journal of Finance	78	78	78	67**	78	78
Journal of International Business Studies	105*	89	89	89	89	89
Journal of Marketing	61	61	61	61	61	61
Review of Economic Studies	69*	50	50	68*	50	50

* Difference caused by treating online first year as year of publication

** Difference caused by treating year of publication as NBER/SSRN working paper as year of publication

Table 4 reports on citation levels for articles published in the 2009 volume of the six journals. Two patterns are apparent here. First of all, just like for my own publication record, Crossref and Dimensions show citation levels that are very similar or even nearly identical to Scopus and are higher than the Web of Science. When compared to Google Scholar citation levels, average citation coverage is 36.5% for Crossref, 38.5% for Dimensions, 38.6% for Scopus, and

33.9% for the Web of Science. At 99.5% Microsoft Academic citation levels are almost identical to Google Scholar.

Table 4: Citation coverage for 2009 volume of six top journals across six data sources

Journal	Crossref	Dimensions	Google Scholar	Microsoft Academic	Scopus	Web of Science
Academy of Management Journal	9368	9731	21234	20917	10059	8746
% of Google Scholar citations	44%	46%	100%	99%	47%	41%
Journal of Accounting Research	2210	2330	7187	7032	2282	2002
% of Google Scholar citations	31%	32%	100%	98%	32%	28%
Journal of Finance	8423	8868	28314	22254**	8687	7618
% of Google Scholar citations	30%	31%	100%	N/A	31%	27%
Journal of International Business Studies	9109*	8291	17176	17391	8307	7319
% of Google Scholar citations	N/A	48%	100%	101%	48%	43%
Journal of Marketing	9613	10235	23249	23364	10034	8727
% of Google Scholar citations	41%	44%	100%	100%	43%	38%
Review of Economic Studies	3451*	2817	9624	11621*	2924	2634
% of Google Scholar citations	N/A	29%	100%	N/A	30%	27%
Overall coverage as % of GS	36.5%	38.5%	100%	99.5%	38.6%	33.9%

* Not comparable as articles included are different because online first year is treated as year of publication

** Not comparable as articles included are different because year of publication as NBER/SSRN working paper is treated as year of publication

Second, when compared to Google Scholar and Microsoft Academic, the citation coverage in Crossref, Dimensions, Scopus and the Web of Science is substantially lower for the three journals in Accounting, Finance and Economics, on average around 30%, than for the three journals in Management, International Business and Marketing, on average around 44%. Obviously, we would need a bigger sample of journals to draw any firm conclusions on this, but it appears as if, even within a single discipline, there is strongly variance in coverage between the different data sources.

Conclusion

Our comparison of coverage across six data sources showed that the two new kids on the block, Crossref and Dimensions, hold their own when compared to Scopus and the Web of Science, but are beaten by Google Scholar and Microsoft Academic. This means that if our findings with regard to Crossref and Dimensions can be confirmed by larger-scale studies, our options for literature search and citation analysis would certainly have widened further.

All of the four free data sources have an edge in terms of recency as they often find publications within weeks of them being published online. In contrast, although Scopus does list “in-press” articles, it typically does so much later than the four other data sources; the Web of Science only enters articles in their database as part of a print publication issue. Given that, especially in the Social Sciences, articles can be available in online first for 1-2 years before being allocated to a print issue, this could be a major drawback when doing a literature review. It is also problematic when reviewing an academic’s publication profile as knowing about their recent publications is important in recruitment and promotion, when searching for reviewers, keynote speakers, and examiners, and even in preparing for a meeting with the academic in question.

The six data sources have very different search options and limitations. For author and journal searches, the ease of disambiguation varies substantially between data sources. For

keyword searches, some of the data sources allow searching for title words only, for others the standard search is in the title, abstract and keywords, yet others allow for full-text searching. The ease of searching for multiple authors or journals in a single search also differs by data source, as do the availability of affiliation searches and the option to *exclude* author names or keywords. Thus, access to six sources for publication and citation data – four of which offering free access – with roughly equivalent publication coverage and varying levels of citation coverage offers academics with a wide array of choices for literature reviews and citation analysis.

References

- Fairhurst, V. (2019). The first year of the Crossref Ambassador Program: highlights and challenges. *Science Editing*, 6(1): 58-63.
- Harzing, A. W. (2007) Publish or Perish available from <https://harzing.com/resources/publish-or-perish>
- Harzing, A.W. (2016). Microsoft Academic (Search): a Phoenix arisen from the ashes?. *Scientometrics*, 108(3): 1637-1647.
- Harzing, A.W., Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison, *Scientometrics*, 106(2): 787-804.
- Harzing, A.W., Alakangas, S. (2017a). Microsoft Academic: is the phoenix getting wings?. *Scientometrics*, 110(1): 371-383
- Harzing, A.W., Alakangas, S. (2017b). Microsoft Academic is one year old: the Phoenix is ready to leave the nest. *Scientometrics*, 112(3): 1887-1894.
- Hug, S.E., Ochsner, M., & Brandle, M.P. (2017). Citation analysis with microsoft academic. *Scientometrics*, 111(1): 371-378.
- Hug, S.E., Brandle, M.P. (2017). The coverage of Microsoft Academic: analyzing the publication output of a university. *Scientometrics*, 113(3): 1551-1571.
- Martin-Martin, A., Orduna-Malea, E., Thelwall, M., & Delgado-López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4): 1160-1177.
- Orduña-Malea, E., Delgado-López-Cózar, E. (2018). Dimensions: re-discovering the ecosystem of scientific information. *El Profesional de la Información*, 27(2): 420-431.
- Thelwall, M. (2017). Microsoft Academic: A multidisciplinary comparison of citation counts with Scopus and Mendeley for 29 journals. *Journal of Informetrics*, 11(4): 201-212.
- Thelwall, M. (2018a). Microsoft Academic automatic document searches: Accuracy for journal articles and suitability for citation analysis. *Journal of Informetrics*, 12(1): 1-9.
- Thelwall, M. (2018b). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, 12(2): 430-435.

ⁱ Our Publish or Perish Dimensions searches for the three names we tested (Eugene Garfield, Blaise Cronin and Mike Thelwall) produced citation levels that were (slightly) above those reported for Scopus in this article. The lack of coverage for author searches in Dimensions reported in this article might thus have been caused by a problem in the author disambiguation system that has since been resolved.

ⁱⁱ Our focus on the Social Sciences might limit generalization as the two commercial databases are less comprehensive for the Social Sciences, Arts & Humanities and to a lesser extent Engineering than for the Sciences and Life Sciences (for a detailed analysis, see Harzing & Alakangas, 2016, 2017a). Hence, differences in coverage between the six data sources might be smaller for the Sciences and Life Sciences.